

# Galaxy Statistics in Pencil-beam Surveys at High Redshifts

Joseph A. Muñoz<sup>\*</sup>, Hy Trac, and Abraham Loeb

*Harvard-Smithsonian Center for Astrophysics, 60 Garden St., MS 10, Cambridge, MA 02138, USA*

13 January 2010

## ABSTRACT

Surveys of faint galaxies at high redshifts often result in a “pencil-beam” geometry that is much longer along the line-of-sight than across the sky. We explore the effects of this geometry on the abundance and clustering of Lyman-break galaxies (LBGs) and Lyman- $\alpha$  emitters (LAEs) in current and future surveys based on cosmological N-body simulations which adequately describe the nonlinear growth of structure on small scales and compare to linear theory. We find that the probability distribution of the LBG abundance is skewed toward low values since the narrow transverse dimension of the survey is more likely to probe underdense regions. Over a range that spans 1–2 orders of magnitude in galaxy luminosities, the variance in the number of objects differs from the commonly used analytic prediction and is not dominated by Poisson noise. Additionally, nonlinear bias on small scales results in a one-dimensional power spectrum of LAEs using a James Webb Space Telescope field-of-view that is relatively flat, markedly different from the expectation of linear perturbation theory. We discuss how these results may affect attempts to measure the UV background at high redshifts, estimate the relationship between halo mass and galaxy luminosity, and probe reionization by measuring the power-modulating effect of ionized regions.

**Key words:** cosmology: theory – cosmology: observations – early universe – large-scale structure of the universe – galaxies: high-redshift – galaxies: abundances

## 1 INTRODUCTION

Over the past few years, searches for Lyman-break galaxies (LBGs) and Lyman-alpha emitters (LAE) have revealed new populations of young, star-forming objects at redshift beyond  $z = 6$  (e.g. Stark et al. 2007a; Richard et al. 2008; Bouwens et al. 2008, 2009) that shed light on the star formation history at high redshifts and the sources responsible for cosmic reionization. In addition, evidence has been found for a population of massive, evolved systems around  $z = 5$  (Mobasher et al. 2005; Wiklind et al. 2008) that hints at an earlier period of higher star formation than has yet to be seen in surveys of bright LBGs and LAEs.

Theoretically, it is expected that the bulk of the star formation during reionization had taken place in less luminous galaxies than previously observed (e.g. Barkana & Loeb 2001; Wyithe & Loeb 2006). Dwarf galaxies are supposed to supply the bulk of the radiation that reionizes the Universe as well as provide the building blocks of the Milky Way and the possible birth place of globular clusters (e.g. Kravtsov & Gnedin 2005; Gnedin & Kravtsov 2006; Moore et al. 2006; Madau et al. 2008; Muñoz et al. 2009). Attempts to probe the faint end of the luminosity function have often sacrificed field-of-view for the sake of higher flux

sensitivity. Objects in the 11.2 arcmin<sup>2</sup> Hubble Ultra Deep Field (HUDF), for example, can be seen in the  $z_{850,AB}$  pass-band down to an apparent magnitude of about 29. In an even more extreme case, a single long-slit spectroscopic survey for gravitationally-lensed LAEs has a field-of-view of only about 32 arcsec<sup>2</sup> (which is even smaller in the unlensed source plane) but can achieve large boosts in sensitivity if positioned on the gravitational lensing critical line of a foreground galaxy cluster. The James Webb Space Telescope (JWST<sup>1</sup>), planned for launch in 2013, will have a better sensitivity, enabling observers to probe “pencil beams” out to even higher redshifts.

Recent data analysis from deep and narrow galaxy surveys has begun to be incorporated into theoretical models of structure formation and reionization. Several studies have attempted to relate the observed UV and Lyman-alpha luminosities from these sources to the masses of the dark matter halos that host them by fitting results from simulations and semi-analytic prescriptions to the observations and so understand the clustering and abundance of such galaxies (e.g. Le Delliou et al. 2005, 2006; Stark et al. 2007b; Kobayashi et al. 2007; Nagamine et al. 2008; Orsi et al. 2008; Kobayashi et al. 2009). In addition, it

<sup>\*</sup> E-mail: jamunoz@cfa.harvard.edu

<sup>1</sup> <http://www.jwst.nasa.gov/>

has been shown that cosmic reionization in the high-redshift Universe can be probed by observing its effects on the galaxy power spectrum (e.g. Babich & Loeb 2006; McQuinn et al. 2007; Wyithe & Loeb 2007).

In this paper, we study the underlying abundance and clustering statistics of halos in narrow fields-of-view. Our goal is to relate the theory of structure formation to the large amount of observational data from high-redshift galaxy surveys. We take into account the effect of large-scale density fluctuations as well as the nonlinearities on small scales. We use state-of-the-art numerical simulations to examine the field-to-field sample distribution and one-dimensional power spectrum of “pencil-beam” galaxy surveys that are much longer along the line-of-sight than across the field-of-view. Our analysis is designed to help observers in the interpretation of related observational data in the future.

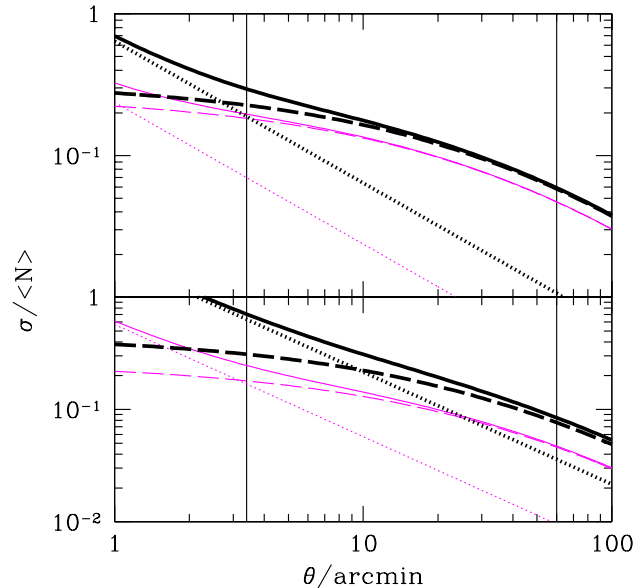
While, similar studies of the cosmic variance and count distributions in high-redshift dropout surveys were carried out (e.g. Trenti & Stiavelli 2008; Overzier et al. 2009), our analysis differs from Trenti & Stiavelli (2008) in that we use our simulated halo catalogs to create mock surveys viewed along the light-cone and probes more deeply than both studies since our simulations can resolve halos as small as  $\sim 10^8 M_\odot$ .

Our paper is organized as follows. In §2 we describe our simulations. The probability distributions of object counts in high-redshift pencil surveys are analyzed in §3. In §4, we explore the one-dimensional power spectrum obtained from such surveys, the nonlinear bias, and the usefulness of such measures in relating galaxy luminosity and halo mass. Finally, we discuss our results and their implications for observations in §5.

## 2 THE SIMULATION DATA

To simulate observations of high-redshift galaxies, we adopt halo catalogs from an N-body simulation and semi-analytically dress each halo with a luminosity based on its mass. These catalogs were produced by the simulation used in Trac et al. (2008) which resolved halos with masses as low as  $\sim 10^8 M_\odot$ , almost two orders-of-magnitude smaller than those resolved in simulations by Trenti & Stiavelli (2008), in a box of comoving size 143 Mpc (assuming a Hubble constant of  $H_0 = 70 \text{ km/s/Mpc}$ ). This approach allows us to probe the faintest high-redshift galaxies within a sufficiently large simulation box that includes the important large-scale fluctuations. A flat,  $\Lambda$ CDM cosmology was assumed with cosmological parameters from the *Wilkinson Microwave Anisotropy Probe* 5-year data release (Komatsu et al. 2009; Dunkley et al. 2009),  $(\Omega_m, \Omega_\Lambda, \Omega_b, h, \sigma_8, n_s) = (0.28, 0.72, 0.046, 0.70, 0.82, 0.96)$ .

We determined a UV or Lyman- $\alpha$  luminosity for each halo in the simulation according to the prescription of Stark et al. (2007b) using their simple assumption that each halo hosts no more than a single bright galaxy. Their model for star formation assumes a fraction,  $f_*$ , of the baryons in a halo with mean baryon fraction forms stars in an amount of time given by the cosmic age at the observed redshift times the duty cycle  $\epsilon_{DC}$ . The efficiency is assumed to be constant with mass for halos with circular velocity above  $100 \text{ km s}^{-1}$ , however, star formation in smaller halos is suppressed so



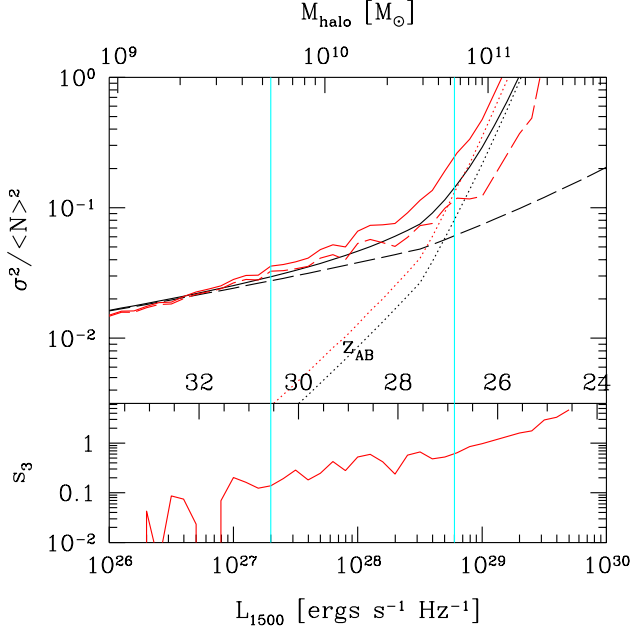
**Figure 1.** The contributions to the total variance (Eq. 1; solid lines) in LBG dropout surveys as a sum of cosmic variance (dashed lines) and Poisson shot noise (dotted lines) contributions (i.e. the first and second terms, respectively, on the right-hand-side of Eq. 1). The top and bottom panels show results for surveys extending from  $z=6-8$  and  $z=8-10$ , respectively. Thin lines assume a luminosity threshold of  $z_{850,AB}=29$ , while for thick ones, the cut is at  $z_{850,AB}=27$ .

that the efficiency is a power-law with mass to the  $2/3$  power. The resulting star formation rate is converted to a UV luminosity (for LBGs) according to the Madau et al. (1998) model or to a Lyman- $\alpha$  luminosity (for LAEs) by assuming a fixed fraction,  $T_\alpha$ , of the  $4 \times 10^{53}$  ionizing photons emitted per star formation rate in  $M_\odot/\text{yr}$  (for a Salpeter IMF) escape the galaxy and the IGM. For a particular type of galaxy, luminosity  $L$  is associated with halo mass  $M_{\text{halo}}$  so that the abundance of observed systems with luminosity  $L$  is equal to the Sheth & Tormen (1999) abundance of halos with mass  $M$  times the duty cycle. We adopt the Stark et al. (2007b) best-fit values of  $(\epsilon_{DC}, f_*) = (0.25, 0.16)$  and  $(1.0, 0.063)$  for LBGs and LAEs, respectively. Since  $\epsilon_{DC} < 1$  for LBGs, we appropriately consider only a fraction,  $\epsilon_{DC}$ , of them to be observed in each snapshot of a galaxy survey.

Finally, since we hope to present results that are not too model-dependent, we will continue to refer to both the galaxy luminosity and halo mass in the calculations that follow and in relevant figures. In this way, future models deemed more appropriate may be applied to our results for a given halo mass.

## 3 PROBABILITY DISTRIBUTION OF GALAXY COUNTS

In this section, we create mock samples of LBGs in pencil beams surveys at different redshifts. Since photometric redshifts of LBGs are typically imprecise, we consider groups of these galaxies in the redshift intervals of  $z = 6-8$  and  $z = 8-10$ . Many LBGs already have been observed in the



**Figure 2.** The upper panel shows relative contributions to the fractional variance in the number counts of galaxies as a function of UV luminosity, z-band magnitude, or host halo mass in counts of LBGs in a dropout survey spanning the redshift interval  $z = 6-8$  with a  $3.4' \times 3.4'$  field-of-view. Solid lines show the total variance, while long-dashed and dotted lines show the contributions from sample variance and Poisson noise, respectively. Red curves show simulation results, while black curves were calculated analytically. Vertical lines bracket the region where the variance is higher than expected due to the skewness of the full count probability distribution but is not Poisson-dominated. The lower panel shows the skewness of the full galaxy count probability distribution calculated from the simulation (Eq. 4).

HUDF around  $z = 6$  (e.g. Bouwens & Illingworth 2006; Bouwens et al. 2006), but observers have not yet had the same success at higher redshifts (e.g. Bouwens et al. 2008, 2009). Given its higher sensitivity, it is hoped that JWST will be able to observe LBGs at these redshifts.

To construct our samples of LBGs, we stitch together simulation slices in the given redshift range so that each survey volume is viewed along the light-cone. This allows the mass function of halos to evolve along the line-of-sight, as needed to represent high-redshift surveys and to correctly analyze their statistics (Muñoz & Loeb 2008). For simplicity, we have assumed a top-hat selection function over each redshift interval.

The time spacing between simulation snapshots is 10 Myr. This means that there are 29 and 16 simulation slices between  $z = 6-8$  and  $z = 8-10$ , respectively. The comoving sizes of these slices vary from 21.1 Mpc at  $z = 6$  to 27.3 Mpc at  $z = 8$  and 32.6 Mpc at  $z = 10$ . We include in our  $z = 6-8$  lightcone catalog halos from a slice 21 Mpc thick at one end of the simulation snapshot at  $z = 6$ , those from the adjacent slice 21 Mpc thick in the snapshot taken 10 Myr earlier, and so on until we include the halos in a slice 27 Mpc thick at  $z = 8$ . The  $z = 8-10$  lightcone catalog is produced similarly.

The comoving distances within the redshift interval  $z = 6-8$  and  $z = 8-10$  are about 700 and 500 Mpc,

respectively, while the simulation box is only 143 Mpc on a side. In the process of stitching together slices, when we come to the end of the simulation box, we simply rotate the box and again stitching together slices through the box. We make the rotation by giving new coordinates  $(x'_1, x'_2, x'_3)$  to each halo with previous position  $(x_1, x_2, x_3)$  such that  $(x'_1, x'_2, x'_3) = (x_2, x_3, x_1)$ ; we also make the additional rotation  $(x'_1, x'_2, x'_3) = (x_1, x_3, x_2)$  every third time we run through the simulation box so that there is no repetition in skewers along the line-of-sight. These rotations suppress power in the survey on the scale of the simulation length.

For each redshift range, we calculate the field-to-field variance in the count of LBGs in a “pencil beam” of a given field-of-view using both linear theory and from the simulation data. The total variance in a survey is the sum of contributions from cosmic variance and Poisson shot noise (see, e.g., Somerville et al. 2004):

$$\frac{\sigma_{\text{tot}}^2}{\langle N \rangle^2} = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle^2} = \frac{\sigma_{\text{hh}}^2}{\langle N \rangle^2} + \frac{1}{\langle N \rangle}, \quad (1)$$

where  $\langle N \rangle$  is the mean number of LBGs in each skewer. The cosmic (or sample) variance (the first term on the right hand side of Eq. 1) results from the field sometimes lying in a region of high density with a lot of structure due to the clustering of galaxies and sometime lying in an under-dense region or a void. This contribution can be calculated from linear perturbation theory (i.e. based on the linear power-spectrum,  $P(k)$ ) as a function of the minimum halo mass,  $M$ , hosting observed galaxies as

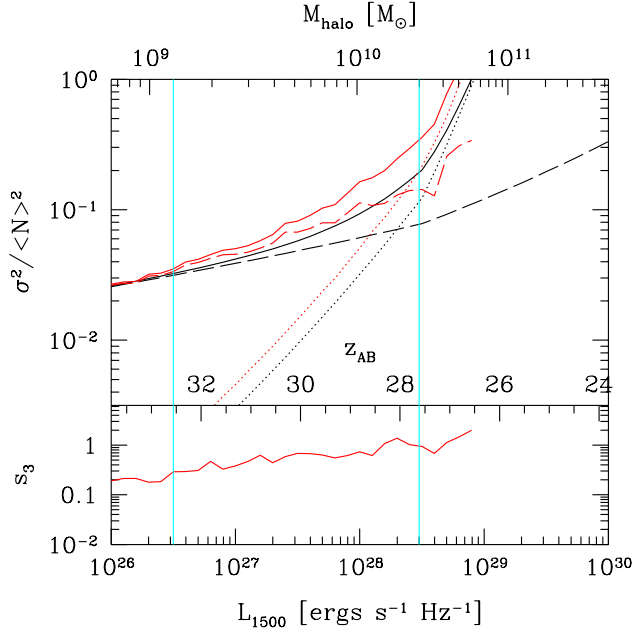
$$\sigma_{\text{hh}}^2(M) = \frac{(b_{\text{eff}}(M) D(z))^2}{(2\pi)^3} \int P(k) W_{xyz}^2 d^3\vec{k}, \quad (2)$$

where  $W_{xyz} = W(k_x) W(k_y) W(k_z)$ ,  $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ ,  $D(z)$  is the linear growth factor evaluated, for simplicity, at the midpoint of the given redshift range, and  $b_{\text{eff}}$  is the Sheth et al. (2001) linear bias integrated over all masses above  $M$  and weighted by the (Sheth & Tormen 1999, hereafter ST) halo mass function (Matarrese et al. 1997; Muñoz & Loeb 2008). The window function,  $W(k_i)$ , is the Fourier transform of a top-hat in the  $i$ -th dimension and is given by:

$$W(k_i) = \frac{\sin(k_i a_i/2)}{k_i a_i/2}. \quad (3)$$

In equation (3),  $a_x$  and  $a_y$  are the narrow dimensions of the skewer-shaped survey evaluated at  $z$ . The subscript  $w$  refers to coordinates along the line-of-sight. The length of the survey,  $a_w$ , is given by the comoving distance between the redshift range of interest.

To calculate the survey statistics, we take  $\theta = 3.4'$  (the field-of-view the HUDF and approximately of JWST) and count the number of LBGs in each of the 256 (225) possible skewer positions perpendicular to the front face of our  $z = 6-8$  ( $z = 8-10$ ) lightcone simulation volume. While nearby skewers may be correlated, distant skewers are slightly anticorrelated. Taking each possible skewer exactly once is equivalent to choosing an infinite number of randomly located skewers since each skewer position will be sampled the same number of times on average. Moreover, given that we expect correlations on the size of the simulation box to be negligible, there is no difference between this statistical measurement and one with only a single skewer



**Figure 3.** Same as Figure 2 but for the redshift interval  $z = 8-10$ .

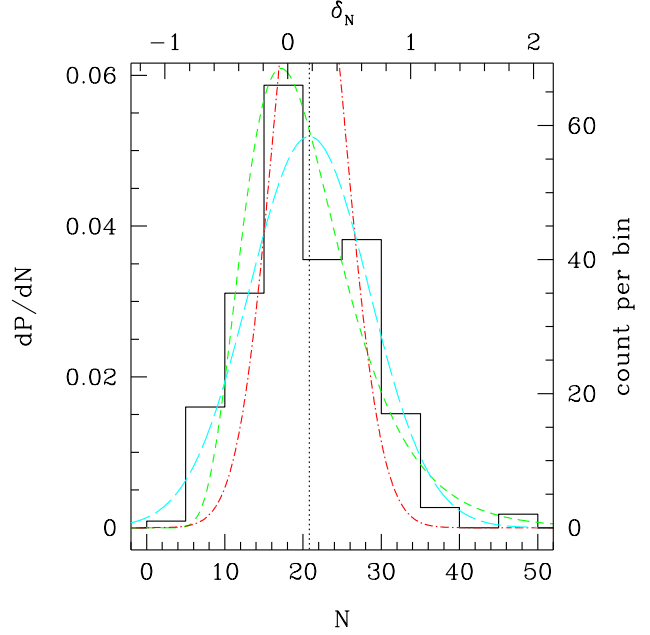
and multiple realizations of the simulation volume are produced using different initial conditions.

Figure 1 compares the contributions from cosmic and Poisson variance as calculated by linear theory (Eq. 1 and 2) for  $a_x = a_y$  as a function of the opening angle of the survey,  $\theta = a_x / \chi(z)$ . This diagram can be used to calculate the effectiveness of future surveys with large fields of view. Given the limiting size of our simulation volume, a sufficient number of skewers is not obtainable to perform comparable numerical calculations for large surveys. However, figures 2 and 3 show the agreement for the HUDF field-of-view.

Figures 2 and 3 compare the contributions to the variance in the simulations with calculations based on linear theory for the  $z = 6-8$  and  $z = 8-10$  ranges, respectively. They suggest that the observed statistics are well approximated by the analytic calculations at the low luminosity limits. However, while linear theory predicts a Gaussian probability distribution of the count of halos with variance given by equation 2, the simulated probability distribution of bright LBGs has a non-Gaussian shape. This can be clearly seen by eye in Figure 4, where the solid histogram representing the measured probability distribution is best represented by a log-normal distribution. The skewness as a function of minimum luminosity is presented in the bottom panels of figures 2 and 3. We define the skewness as the third moment of the probability distribution normalized by the variance to the  $3/2$  power:

$$s_3 = \frac{\langle (N - \langle N \rangle)^3 \rangle}{(\sigma^2)^{3/2}}. \quad (4)$$

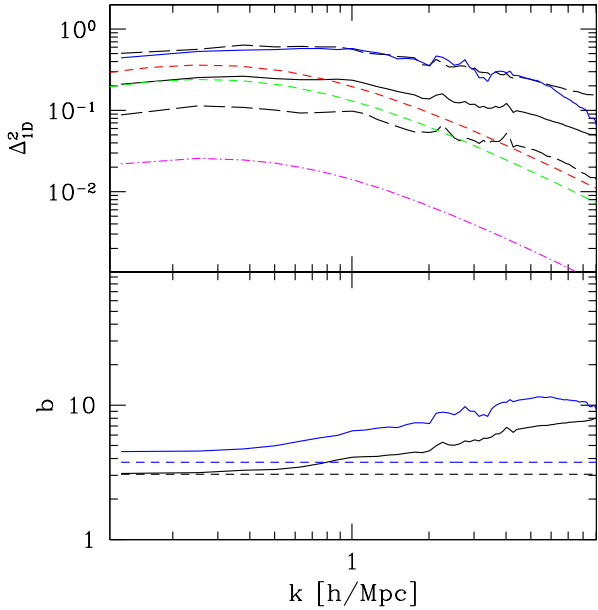
The seemingly large amplitude variations in the skewness at low luminosity for  $z = 6-8$  are due to small numerical fluctuations around the nearly zero skewness from the simulation, plotted on a log scale. Deviations between the analytic and simulation values of the sample variance grow as the skewness becomes more significant. This behavior is



**Figure 4.** The full probability distribution for the number  $N$  of LBGs in a dropout survey spanning the redshift interval  $z = 8-10$  with a minimum galaxy luminosity of  $10^{28}$  ergs/s/Hz. The solid black curve shows the distribution as a histogram measured from the simulation, while the red dot-dashed, cyan long-dashed, and green short-dashed lines overlay Poisson, Gaussian, and log-normal distributions with the same mean, mean and standard deviation, and log-mean and log-standard deviation, respectively. The vertical, dotted line denotes the mean number of objects in the survey over many pointings. The distributions are plotted as differential probability in a differential abundance bin on the left, vertical axis as a function of the measured abundance of LBGs,  $N$ , on the lower, horizontal axis and as a function of the overdensity in the abundance  $\delta_N = N / \langle N \rangle - 1$  on the upper, horizontal axis. Additionally, the counts of pointings per abundance bin for the simulation histogram is shown on the right, vertical axis. These counts per bin show that Poisson errors in the simulation histogram are smaller than the differences between the overlaid distributions. The simulation results match most closely with the skewed log-normal distribution.

a manifestation of nonlinear clustering on the small scales probed by the narrowness of the skewer.

Finally, the figures show sharp regimes where the field-to-field variance is dominated either by cosmic or Poisson variance. For the two redshift ranges  $z = 6-8$  and  $z = 8-10$  this division comes at around  $L(M = 5 \times 10^{10} M_{\odot}) = 6 \times 10^{28}$  ergs/s/Hz and  $L(M = 2 \times 10^{10} M_{\odot}) = 3 \times 10^{28}$  ergs/s/Hz, respectively. The nonlinearity at higher luminosities causes the sample variance to dominate at masses which are  $\sim 50-100\%$  larger than otherwise expected. However, since the Poisson variance is unaffected by nonlinearities, there remains only a specific mass range (shown in the figures) over which the total variance is different than expected.

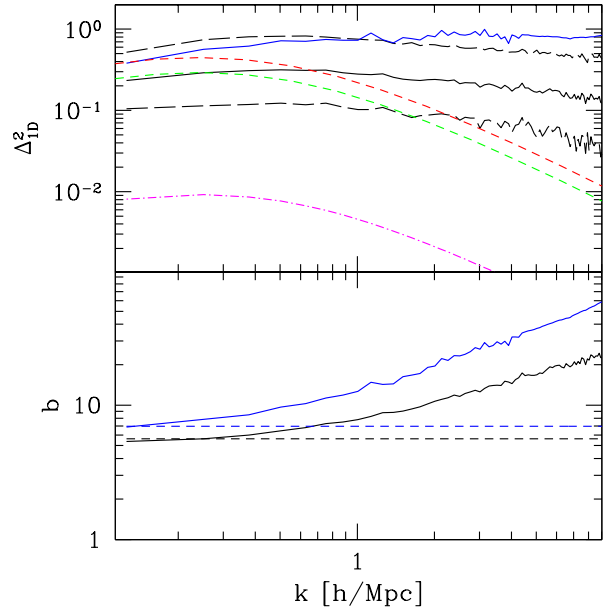


**Figure 5.** The 1D power spectrum of halos hosting LAEs in a survey with a  $3.4' \times 3.4'$  field-of-view at  $z = 6$ . The black (blue) lines represent values for a minimum Lyman- $\alpha$  luminosity of  $L_{\text{Ly}\alpha, \text{min}} = 10^{40}$  ( $10^{41}$ ) ergs/s corresponding to  $M_{\text{halo}} = 1.6 \times 10^9$  ( $6.4 \times 10^9$ )  $M_{\odot}$ . In the upper panel, solid lines show the average log of the Poisson-subtracted, dimensionless power spectrum. Long-dashed lines indicate the estimated  $1 - \sigma$  variation in the log of the amplitude of the power spectrum from skewer-to-skewer. For clarity, these lines only appear for the lower luminosity threshold, but they are about the same size for the higher luminosity one. Short-dashed, green and red lines show the linear theory predictions for the dimensionless halo power spectrum given the lower and higher luminosity thresholds, respectively, while the dotted-dashed, magenta curve shows the linear theory calculations for the dark matter. The lower panel shows the non-linear bias (Eq. 8) measured from the simulation (solid lines) and the linear, ST bias (short-dashed lines).

#### 4 LINE-OF-SIGHT (ONE-DIMENSIONAL) POWER SPECTRUM

Next, we consider the one-dimensional (1D) galaxy power spectrum measured along the line-of-sight of a “pencil-beam” survey at high redshift with the same  $3.4' \times 3.4'$  field-of-view considered in the previous section. We explore what the information along the radial direction can or cannot tell us about the nature of high-redshift galaxies given the narrow field-of-view. We also examine whether or not it can be used to distinguish between luminous galaxies hosted by different mass halos, and ultimately to probe reionization. We are considering the “galaxy” power spectrum in the sense that we apply a duty cycle for LAEs and translate halo masses into hosted Lyman- $\alpha$  luminosities, but our results do not include the effects of reionization on the power spectrum that we expect to observe in real surveys (e.g. Babich & Loeb 2006; McQuinn et al. 2007; Wyithe & Loeb 2007).

When analyzing the 1D power spectrum, we do not attempt to produce mock surveys viewed along the light-cone since we wish to study the bias of the halos hosting high-



**Figure 6.** Same as Figure 5 except at  $z = 10$ . The minimum Lyman- $\alpha$  luminosity of  $L_{\text{Ly}\alpha, \text{min}} = 10^{40}$  ( $10^{41}$ ) ergs/s corresponds to  $M_{\text{halo}} = 8.2 \times 10^8$  ( $3.2 \times 10^9$ )  $M_{\odot}$  at this redshift.

redshift galaxies and compare our results to analytic, linear theory. Rather, we consider snapshots at  $z = 6$  and  $z = 10$ . For this purpose, it is preferable that the evolution of the halo mass and correlation functions will not introduce features in the power spectrum (Muñoz & Loeb 2008). Additionally, stitching together simulation slices to produce a view along the light-cone would unnecessarily add or subtract (depending on the prescription used) power from the spectrum on scales larger than the size of the simulation box and on the scale of the separation between simulation slices. As shown in Figures 5 and 6, the 143 Mpc comoving size of the simulation box is sufficiently large to reproduce the linear results on the largest scales. We plot our results only for values of  $k$  larger than  $0.1 \text{ Mpc}^{-1}$  to avoid confusion from power loss due to the finite box size.

We calculate the spectrum in terms of real-space wavenumbers, as opposed to ones in redshift-space, and ignore contamination of distance measurements by peculiar velocities (Monaco et al. 2005) since they are expected to be small in the early universe. The redshifts of LAEs can be constrained to within a few hundred km/s, corresponding to a wave-number of  $k \sim 3 \text{ Mpc}^{-1}$ , while the distances to high-redshift LBGs can be uncertain to hundreds of comoving Mpc (Bouwens & Illingworth 2006).

We compare the power spectrum calculated from the simulation using a 1D Fast Fourier Transform (FFT) to that determined analytically from linear perturbation theory. The theoretical 1D power spectrum is obtained by integrating the 3-dimensional power spectrum over a window function corresponding to the field-of-view (Kaiser & Peacock 1991):

$$P_{1D}(k_w, M) = \frac{(b_{eff}(M) D(z))^2}{(2\pi)^2} \int P(k) W_{xy}^2 dk_x dk_y, \quad (5)$$

where  $M$  is the minimum halo mass of the LBGs under con-

sideration,  $k = \sqrt{k_x^2 + k_y^2 + k_w^2}$ ,  $W_{xy} = W(k_x)W(k_y)$ , and  $W(k_i)$  is the same as in equation 3. The power measured in the simulation, however, includes an additional, scale-independent component from Poisson noise. This, too, can be calculated analytically as

$$P_{1D, \text{Poisson}}(M) = (\bar{n}(> M) a_x a_y a_w Q(a_w))^{-1} \quad (6)$$

and subtracted off, where  $\bar{n}(> M)$  is the ST halo mass function of LAE hosts with mass greater than  $M$  and

$$Q(a_w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W^2(k_w) dk_w. \quad (7)$$

Here,  $a_z$  can be any length scale and represents the depth of the skewer for which  $\sigma_P^2 = \langle N(> M) \rangle = \bar{n}(> M) a_x a_y a_w$  is the Poisson variance in the number of LAEs. Since,

$$\int_{-\infty}^{\infty} W^2(k_w) dk_w \sim 2/a_w,$$

$Q(a_w) \propto 1/a_w$ , and so  $P_{1D, \text{Poisson}}(M)$  is approximately independent of  $a_w$ , as expected. From equation 5, we can see that the 1D power spectrum at a given scale is a complicated convolution of shorter wavelength modes with the window function. This integration washes out any features that may have existed at small scales. Thus, we expect the 1D spectrum to be relatively featureless and smooth.

We analyze the power spectrum for as many skewers through our simulation box that can be packed into the 143 Mpc size,  $17^2$  and  $15^2$  for  $z = 6$  and  $z = 10$ , respectively, before rotating the box to a new orientation and collecting additional sets of skewers. As in §3, this scheme is equivalent to choosing an infinite number of skewers each placed randomly on the face of the simulation box. The argument also applies to additional skewers gained through rotation of the simulation volume.

Figures 5 and 6 compare analytic and simulation calculations of the 1D power spectrum and bias after cosmic reionization is assumed to be completed ( $z = 6$ ) and in its early stages at  $z = 10$ , respectively. In both figures, the black (blue) lines represent values for a minimum Lyman- $\alpha$  luminosity of  $L_{L\alpha, \min} = 10^{40}$  ( $10^{41}$ ) ergs/s corresponding to  $M_{\text{halo}} = 1.6 \times 10^9$  ( $6.4 \times 10^9$ )  $M_\odot$  at  $z = 6$  and  $M_{\text{halo}} = 8.2 \times 10^8$  ( $3.2 \times 10^9$ )  $M_\odot$  at  $z = 10$ . While LAEs are currently detected to approximately this luminosity at  $z = 6$ , future surveys should be able to probe down to the same luminosity at  $z = 10$ . Solid lines show the average log of the Poisson subtracted power spectrum (represented as  $\Delta_{1D}^2(k) = k P_{1D}(k)/\pi$ ) measured from  $3 \times 17^2$  and  $3 \times 15^2$  skewers for  $z = 6$  and  $z = 10$ , respectively. The factor of 3 comes from additional independent skewers we obtained from rotating the simulation box.

Long-dashed lines indicate the estimated  $1 - \sigma$  variation in the log of the amplitude of the power spectrum from skewer-to-skewer. For clarity, these lines are only shown for  $L_{L\alpha, \min} = 10^{40}$  ergs/s, but they are about the same size for  $L_{L\alpha, \min} = 10^{41}$  ergs/s. An estimate of the power from a single skewer will fall within the  $1 - \sigma$  bounds 68 percent of the time. The standard deviation of the log of the power is approximately the same for all redshift and luminosity threshold combinations we considered with  $\sigma \approx 0.4$ ; it is also approximately independent of scale. The precision to which one is able to recover the true mean power spectrum is given by the standard error:  $\sigma_e = \sigma/\sqrt{N_{\text{skewers}}}$ . Assuming

that our estimate of the standard deviation is close to the true value and given our number of independent skewers, we were able to estimate the mean of the log of the power to within  $\sigma_e \approx 0.02$ .

The level of error in measuring the power informs whether the 1D power contains sufficient information to distinguish between galaxies of different masses and luminosities. This differentiability is important in constraining the mass-luminosity relationship and happens when the standard errors for each power spectrum using  $N_{\text{skewers}}$  independent fields-of-view (in this case all having dimensions of  $3.4' \times 3.4'$ ) are small enough that they don't overlap:  $|\mu_{41} - \mu_{40}| = (\sigma_{41} + \sigma_{40})/\sqrt{N_{\text{skewers}}}$ , where  $\mu_x$  and  $\sigma_x$  are the mean and standard deviations of the log of the power for the luminosity thresholds corresponding to  $L_{L\alpha, \min} = 10^x$  ergs/s. On the largest scales probed, the mean of the log of the power differs by  $\approx 0.23$  for both redshifts (with a weak scale dependence) and  $\sigma_{41} \approx \sigma_{40} \approx 0.4$ . Thus, approximately 12 independent skewers are needed to distinguish between the two luminosity thresholds based on their 1D power spectrum.

The short-dashed green (red) line shows the analytic power-spectrum according to linear perturbation theory, with the same luminosity threshold as the black (blue) line. With no fit other than the calculations described above, the analytic predictions match the data from the simulations quite well on the largest scales. As soon as nonlinearities dominate on small scales, the simulation results and the analytic calculations diverge in their predictions. However, it is important to note that the nonlinearities at work are on even smaller scales than the deviation scale due to aliasing.

In the bottom panels of Figures 5 and 6, we show the halo bias,

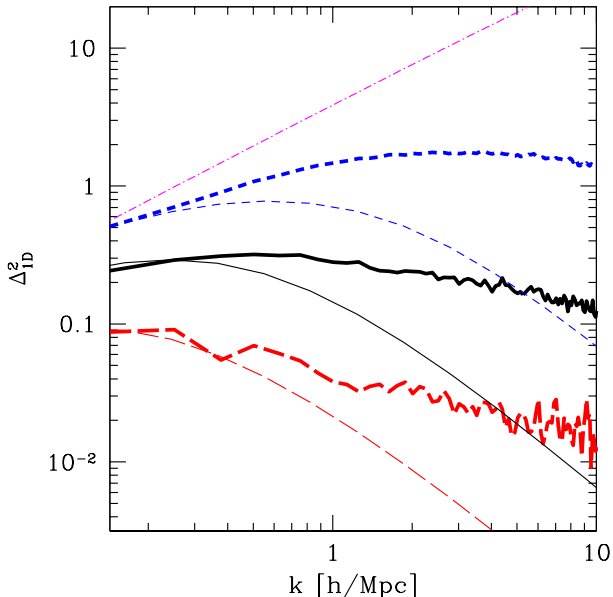
$$b = \sqrt{P_{1D, \text{halo}}^{\text{halo}}/P_{1D, \text{dm}}^{\text{analytic}}}, \quad (8)$$

measured from the simulation (solid lines) and compare the results to the analytic ST calculation (short-dashed lines). Here,  $P_{1D, \text{dm}}^{\text{analytic}}$  is the linear, analytic calculation of the 1D dark matter power from equation (5) with  $b_{\text{eff}}(M) = 1$ . These plots show the same trend as in the upper panels with increasing divergence between simulation and analytic calculations on small scales. This is precisely the result obtained from the full three-dimensional power by Trac & Cen (2007). However, note that the linear and nonlinear values do not match up exactly for  $L_{L\alpha, \min} = 10^{41}$  ergs/s at  $z = 6$ . The ST mass function is not a perfect fit over all masses and so it results in a slight deviation in the derived bias. This effect also manifests itself in the upper panel as a small deviation on large scales between the simulation and analytic power for the same luminosity and redshift.

A striking aspect of the 1D power spectrum illustrated in the top panels is that, for the  $3.4' \times 3.4'$  field-of-view, it remains relatively flat and featureless over a range of  $k$ -values spanning an order-of-magnitude or more. This smoothness is expected since both the aliasing of small-scale power and the averaging of many skewers wash out small features in the spectrum. However, the flatness is unique to our choice of field-of-view.

In Figure 7, we investigate how observational results from different fields-of-view can be compared by plotting the 1D power simulated in  $3.4' \times 3.4'$  pencil-beam skewers





**Figure 7.** A comparison of the 1D power spectrum of halos hosting LAEs, as in Fig. 6, in surveys with different fields-of-view at  $z = 10$ . We set  $L_{\text{LA},\text{min}} = 10^{40}$  ergs/s for each survey. Thick lines show the simulated average log of the Poisson-subtracted, dimensionless power spectrum, while thin ones represent calculations from linear theory. Solid, black lines correspond to a  $3.4' \times 3.4'$  field-of-view, while the short-dashed, blue and long-dashed, red lines denote  $1' \times 1'$  and  $10' \times 10'$  fields-of-view, respectively. The field-to-field variance from Figures 5 and 6 have been omitted for clarity. For reference, we have plotted the linear theory spectrum for the dark matter assuming a pure 1D survey with an infinitesimal field-of-view as the dotted-dashed, magenta curve.

alongside those using  $1' \times 1'$  and  $10' \times 10'$  fields-of-view. The behavior in the interesting range of modes differs among the three cases because of the different modes allowed in the integration by the window function, which mixes up modes in a way that cannot be easily de-convolved (see Eq. 5).

Comparing the the linear theory predictions, we see that the position of the turn-over is directly related to the size of the window function with the narrower field-of-view turning over at larger  $k$ . Although the simulated position of the turn-over still increases monotonically with field-of-view size, the aliasing of nonlinear small-scale power changes the window function convolution in such a way that the position of the peak is not easily determined from the linear case.

It is also clear that the flatness we noted in the simulated spectrum for the  $3.4' \times 3.4'$  field-of-view is not generic; the spectrum for the narrowest field-of-view first rises with  $k$  and then decreases. The  $10' \times 10'$  field-of-view produces a spectrum that is flatter than the one produced by the  $1' \times 1'$  field-of-view, but since the peak is also at much larger scales, we end up probing the spectrum in the regime where it begins to fall off at small scales. The extra bumps and wiggles in the curve result from fewer skewers of that size being available for averaging.

For reference, we have also shown in Figure 7 the 1D dark matter power spectrum in an idealized 1D survey with an infinitesimal field-of-view. The aliased power in this case increases the amplitude of the spectrum even more than the

bias factor in the LAE spectra using realistic fields-of-view. The figure shows the 1D power approaching this limit as the field-of-view for realistic surveys decreases.

## 5 CONCLUSIONS

In this work we have analyzed the statistics of galaxies in mock “pencil beam” surveys of high-redshift LBGs and LAEs with narrow fields-of-view produced using N-body simulations. We computed the effect of sample variance and nonlinear bias and showed how these will influence measurements of galaxy counts in existing and upcoming surveys.

While the variance due to fluctuations in the local overdensity in LBG surveys increases almost monotonically with increasing luminosity (or halo mass), it dominates only at low luminosities and is overtaken at high luminosities by a Poisson variance that rises more steeply as the average number of galaxies in the survey drops. However, the overdensities become nonlinear on small scales. A given point is more likely to be in underdense regions than in overdense regions because of their larger volume filling fraction. The resulting skewness of the probability distribution of galaxy counts creates slightly more uncertainty in obtaining the true number density of galaxies at the observed redshift than would otherwise be expected. This can be seen graphically in Fig. 4 where the simulation histogram is most closely matched by a skewed, log-normal distribution whose peak is centered on negative values of the overdensity. The skewness of the distribution and the resulting increased probability of surveying an underdense region should be taken into account when interpreting galaxy counts in the context of the ionizing UV background they produce at high redshift, or else it would result in an underestimate of the background. Precise measurements of the cosmic variance in a survey is also important for understanding the differences between two independent samples. Our results and methods can be used to calibrate these errors for different fields-of-view and survey depths appropriate to a given study instead of relying only on linear theory estimates at a single redshift as in Bouwens et al. (2008).

Whether or not future surveys of LAEs will have the angular resolution to compute angular correlation functions in the plane of the sky, they will probe many more modes along the line-of-sight. We expect the 1D power spectrum from surveys with a JWST-sized field-of-view to exhibit a roughly flat, smooth profile. Aliasing and the averaging of many skewers are expected to result in a relatively featureless spectrum unable to probe void structure like that seen in Broadhurst et al. (1990). The flatness, on the other hand, is unique to the particular choice of field-of-view, and comparing JWST results, such as measurements of the galaxy bias, with those from other surveys will not be trivial due to the complicated convolution with the window function. Nonetheless, if the survey parameters are known, the amplitude on the largest scales matches well with the expectation from linear theory. However, on small scales, the nonlinear bias becomes large. Given their limited fields-of-view, a survey with JWST will need tens of pointings to be able to distinguish between host halo masses different by an order-of-magnitude limiting its usefulness in determining their mass-to-light ratios. Finally, our calculations ignored the complex

effects of reionization on the amplitude of the power spectrum. The mass-to-light ratio for LAEs at redshifts during reionization could first be calibrated by extrapolating determinations made after reionization, through either clustering or abundance matching. With this information in hand, differences between our results and observations will show the effect of Lyman- $\alpha$  transmission through the HII regions around ionizing sources. However, probing reionization through the changing amplitude of the 1D spectrum (McQuinn et al. 2007) will require particular consideration of the field-of-view.

## 6 ACKNOWLEDGEMENTS

We would like to thank Mark Dijkstra for useful discussions. HT is supported by an Institute for Theory and Computation Fellowship. This research was also supported in part by NASA grants NNX08AL43G and LA and by Harvard University funds.

## REFERENCES

- Babich, D., & Loeb, A. 2006, *ApJ*, 640, 1
- Barkana, R., & Loeb, A. 2001, *Physics Reports*, 349, 125
- Bouwens, R., & Illingworth, G. 2006, *New Astronomy Review*, 50, 152
- Bouwens, R., Illingworth, G., Blakeslee, J., & Franx, M. 2006, *ApJ*, 653, 53
- Bouwens, R. J., Illingworth, G. D., Bradley, L. D., Ford, H., Franx, M., Zheng, W., Broadhurst, T., Coe, D., & Jee, M. J. 2009, *ApJ*, 690, 1764
- Bouwens, R. J., et al. 2008, *ApJ*, 686, 230
- Broadhurst, T. J., Ellis, R. S., Koo, D. C., & Szalay, A. S. 1990, *Nat*, 343, 726
- Dunkley, J., Komatsu, E., Nolte, M. R., Spergel, D. N., Larson, D., Hinshaw, G., Page, L., Bennett, C. L., Gold, B., Jarosik, N., Weiland, J. L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., & Wright, E. L. 2009, *ApJS*, 180, 306
- Gnedin, N. Y., & Kravtsov, A. V. 2006, *ApJ*, 645, 1054
- Kaiser, N., & Peacock, J. A. 1991, *ApJ*, 379, 482
- Kobayashi, M. A. R., Totani, T., & Nagashima, M. 2007, *ApJ*, 670, 919
- . 2009, *arXiv:astro-ph/0902.2882*
- Komatsu, E., Dunkley, J., Nolte, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Meyer, S. S., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2009, *ApJS*, 180, 330
- Kravtsov, A. V., & Gnedin, O. Y. 2005, *ApJ*, 623, 650
- Le Delliou, M., Lacey, C., Baugh, C. M., Guiderdoni, B., Bacon, R., Courtois, H., Sousbie, T., & Morris, S. L. 2005, *MNRAS*, 357, L11
- Le Delliou, M., Lacey, C. G., Baugh, C. M., & Morris, S. L. 2006, *MNRAS*, 365, 712
- Madau, P., Kuhlen, M., Diemand, J., Moore, B., Zemp, M., Potter, D., & Stadel, J. 2008, *ApJ*, 689, L41
- Madau, P., Pozzetti, L., & Dickinson, M. 1998, *ApJ*, 498, 106
- Matarrese, S., Coles, P., Lucchin, F., & Moscardini, L. 1997, *MNRAS*, 286, 115
- McQuinn, M., Hernquist, L., Zaldarriaga, M., & Dutta, S. 2007, *MNRAS*, 381, 75
- Mobasher, B., Dickinson, M., Ferguson, H. C., Giavalisco, M., Wiklind, T., Stark, D., Ellis, R. S., Fall, S. M., Grogin, N. A., Moustakas, L. A., Panagia, N., Sosey, M., Stiavelli, M., Bergeron, E., Casertano, S., Ingraham, P., Koekemoer, A., Labbé, I., Livio, M., Rodgers, B., Scarlata, C., Vernet, J., Renzini, A., Rosati, P., Kuntschner, H., Kümmel, M., Walsh, J. R., Chary, R., Eisenhardt, P., Pirzkal, N., & Stern, D. 2005, *ApJ*, 635, 832
- Monaco, P., Møller, P., Fynbo, J. P. U., Weidinger, M., Ledoux, C., & Theuns, T. 2005, *A&A*, 440, 799
- Moore, B., Diemand, J., Madau, P., Zemp, M., & Stadel, J. 2006, *MNRAS*, 368, 563
- Muñoz, J. A., & Loeb, A. 2008, *MNRAS*, 386, 2323
- Muñoz, J. A., Madau, P., Loeb, A., & Diemand, J. 2009, *MNRAS*, 400, 1593
- Nagamine, K., Ouchi, M., Springel, V., & Hernquist, L. 2008, *arXiv:astro-ph/0802.0228*
- Orsi, A., Lacey, C. G., Baugh, C. M., & Infante, L. 2008, *MNRAS*, 391, 1589
- Overzier, R. A., Guo, Q., Kauffmann, G., De Lucia, G., Bouwens, R., & Lemson, G. 2009, *MNRAS*, 394, 577
- Richard, J., Stark, D. P., Ellis, R. S., George, M. R., Egami, E., Kneib, J.-P., & Smith, G. P. 2008, *ApJ*, 685, 705
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, 323, 1
- Sheth, R. K., & Tormen, G. 1999, *MNRAS*, 308, 119
- Somerville, R. S., Lee, K., Ferguson, H. C., Gardner, J. P., Moustakas, L. A., & Giavalisco, M. 2004, *ApJ*, 600, L171
- Stark, D. P., Ellis, R. S., Richard, J., Kneib, J.-P., Smith, G. P., & Santos, M. R. 2007a, *ApJ*, 663, 10
- Stark, D. P., Loeb, A., & Ellis, R. S. 2007b, *ApJ*, 668, 627
- Trac, H., & Cen, R. 2007, *ApJ*, 671, 1
- Trac, H., Cen, R., & Loeb, A. 2008, *ApJ*, 689, L81
- Trenti, M., & Stiavelli, M. 2008, *ApJ*, 676, 767
- Wiklind, T., et al. 2008, *ApJ*, 676, 781
- Wyithe, J. S. B., & Loeb, A. 2006, *Nat*, 441, 322
- . 2007, *MNRAS*, 382, 921